

# Verified numerical computation for semilinear elliptic problems with lack of Lipschitz continuity of the first derivative

Kazuaki Tanaka<sup>1,\*</sup>, Michael Plum<sup>2</sup>, Kouta Sekine<sup>3</sup>, Masahide Kashiwagi<sup>3</sup>, Shin'ichi Oishi<sup>3,4</sup>

<sup>1</sup>*Graduate School of Fundamental Science and Engineering, Waseda University, Japan*

<sup>2</sup>*Institut für Analysis, Karlsruhe Institut für Technologie, Germany*

<sup>3</sup>*Faculty of Science and Engineering, Waseda University, Japan*

<sup>4</sup>*CREST, JST, Japan*

**Abstract.** In this paper, we propose a numerical method for verifying solutions to the semilinear elliptic equation  $-\Delta u = f(u)$  with homogeneous Dirichlet boundary condition. In particular, we consider the case in which the Fréchet derivative of  $f$  is not Lipschitz continuous. A numerical example for a concrete nonlinearity is presented.

*Key words:* computer-assisted proof, elliptic boundary value problem, existence proof, lack of Lipschitz continuity, semilinear problem, verified numerical computation

## 1 Introduction

We are concerned with a verified numerical computation method for the following elliptic problem:

$$\begin{cases} -\Delta u = f(u) & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega, \end{cases} \quad (1)$$

where  $\Omega \subset \mathbb{R}^n$  ( $n = 1, 2, 3$ ) is a given domain (i.e., an open connected set) and  $f : H_0^1(\Omega) \rightarrow L^2(\Omega)$  is a given nonlinear operator. Here, assuming that  $H^1(\Omega)$  denotes the first order  $L^2$ -Sobolev space on  $\Omega$ , we define  $H_0^1(\Omega) := \{u \in H^1(\Omega) : u = 0 \text{ on } \partial\Omega \text{ in the trace sense}\}$  with inner product  $(\cdot, \cdot)_{H_0^1(\Omega)} := (\nabla \cdot, \nabla \cdot)_{L^2(\Omega)}$  and norm  $\|\cdot\|_{H_0^1(\Omega)} := \|\nabla \cdot\|_{L^2(\Omega)}$ . Hereafter, we denote  $V = H_0^1(\Omega)$ , and  $V^* = H^{-1}(\Omega) := (\text{dual of } V)$  with the usual sup-norm. Moreover, the  $L^2$ -inner product is simply denoted by  $(\cdot, \cdot)$  if no confusion arises.

Verified numerical computation methods for differential equations originate from Nakao's [10] and Plum's work [13], and have been further developed by many researchers. Moreover, the applicability of such methods to semilinear elliptic boundary value problems has been investigated (see, e.g., [11, 12, 15, 16, 20]). In their frameworks, (1) is transformed into a suitable operator equation for proving the existence of a solution close to a computed numerical approximation. In this paper, by defining  $\mathcal{F} : V \rightarrow V^*$  as

$$\langle \mathcal{F}(u), v \rangle := (\nabla u, \nabla v) - (f(u), v) \quad \text{for } u, v \in V,$$

we first re-write (1) as

$$\mathcal{F}(u) = 0 \text{ in } V^*, \quad (2)$$

and discuss the verified numerical computation for (2). In other words, we first consider the existence of a weak solution to (1) (a solution to (2) in  $V$ ), and then we discuss its  $H^2$ -regularity if necessary.

In particular, we select  $f(u) = u^p$  ( $1 < p < 2$ ) as a Fréchet differentiable operator, the Fréchet derivative of which is, however, not Lipschitz continuous. We are looking for positive solutions

---

*E-mail address:* \* imahazimari@fuji.waseda.jp

to (1), that is, we consider the following problem:

$$\begin{cases} -\Delta u = u^p & \text{in } \Omega, \\ u > 0 & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega. \end{cases} \quad (3)$$

It is well known that the Fréchet differentiability of  $f$  (or of  $\mathcal{F}$ ) is essential for the existing theories of verified numerical computations for solutions to (1) (see again, e.g., [11, 12, 15, 16, 20]). Moreover, the Lipschitz continuity of the Fréchet derivative of  $f$  has been required in real examples, mainly for obtaining convenient mapping properties of the fixed point operator to be constructed, but also to avoid some technical difficulties in computing integrals (often they are needed with verification), the integrands of which contain  $f'(v_N)$  and  $f(v_N)$  for  $v \in V_N$ ;  $V_N$  is a finite dimensional subspace of  $V$ , e.g., spanned by a finite element basis or a Fourier-Galerkin basis. For example, when we set  $f(u) = u^p$  ( $1 < p < 2$ ) as mentioned above, such integrations are difficult to calculate with high-precision as well as difficult to estimate with verification, since, even for smooth functions  $u$ , the second derivative of  $f(u(\cdot))$  is not bounded near points  $x \in \mathbb{R}^n$  such that  $u(x) = 0$ . Such integrations are required at many points in the verification process, e.g., when we estimate the norm of the residual  $\|\mathcal{F}(\hat{u})\|_{V^*}$  for some approximation  $\hat{u} \in V$  with verification, and when we compute verified bounds for the operator norm of the inverse of  $\mathcal{F}'_{\hat{u}} : V \rightarrow V^*$ , where  $\mathcal{F}'_{\hat{u}}$  is the Fréchet derivative of  $\mathcal{F}$  at  $\hat{u} \in V$ .

In this paper, we apply Plum's theorem [15] (see Theorem 2.1) to the verified numerical computation for a solution to (3) with  $p \in (1, 2)$ . To be precise, we prove the existence of a solution to (3), on the basis of Theorem 2.1, in balls centered around a numerically computed approximate solution, in the sense of both norms  $\|\nabla \cdot\|_{L^2(\Omega)}$  and  $\|\cdot\|_{L^\infty(\Omega)}$ . For this purpose, we first try to obtain a numerical inclusion of a solution to (1) with  $f(u) = |u|^{p-1}u$ , i.e., a solution to

$$\begin{cases} -\Delta u = |u|^{p-1}u & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega. \end{cases} \quad (4)$$

After that, we confirm its positiveness with a simple computation in order to verify a solution to (3) (see Section 4 for a method for confirming the positiveness). We remark that, when the Fréchet derivative of  $f$  is not Lipschitz continuous, a function that replaces the Lipschitz constant of the Fréchet derivative of  $f$  is required to be concretely constructed; such a function will be denoted by  $g$  in Theorem 2.1. We also propose a concrete construction of such a function for the  $u^p$ -nonlinearity in Section 2. A numerical integration method for integrands arising from the  $u^p$ -nonlinearity with  $p \in (1, 2)$  will be proposed in Section 3.

## 2 Verification theorem for elliptic problems

In this section, we apply the method summarized in [14, 15, 16] to a verified numerical computation for solutions to (1). Throughout this paper, the norm bound for the embedding  $V \hookrightarrow L^p(\Omega)$  is denoted by  $C_p$ , i.e.,  $C_p$  is a positive number that satisfies

$$\|u\|_{L^p(\Omega)} \leq C_p \|u\|_V \quad \text{for all } u \in V. \quad (5)$$

Since a concrete upper bound for  $C_p$  is important for the verification theory, a formula that gives such an upper bound for a given bounded domain  $\Omega$  is provided in Corollary A.2.

## 2.1 $H_0^1$ error estimation

We use the following verification theorem for obtaining  $H_0^1$  error estimations for solutions to (2), i.e., weak solutions to (1).

**Theorem 2.1** ([15, 16]). *Let  $\mathcal{F} : V \rightarrow V^*$  be a Fréchet differentiable operator. Suppose that  $\hat{u} \in V$ , and that there exist  $\delta > 0$ ,  $K > 0$ , and a non-decreasing function  $g$  satisfying*

$$\|\mathcal{F}(\hat{u})\|_{V^*} \leq \delta, \quad (6)$$

$$\|u\|_V \leq K \|\mathcal{F}'_{\hat{u}} u\|_{V^*} \quad \text{for all } u \in V, \quad (7)$$

$$\|\mathcal{F}'_{\hat{u}+u} - \mathcal{F}'_{\hat{u}}\|_{B(V, V^*)} \leq g(\|u\|_V) \quad \text{for all } u \in V, \quad (8)$$

and

$$g(t) \rightarrow 0 \quad \text{as } t \rightarrow 0. \quad (9)$$

Moreover, suppose that some  $\alpha > 0$  exists such that

$$\delta \leq \frac{\alpha}{K} - G(\alpha) \quad \text{and} \quad Kg(\alpha) < 1,$$

where  $G(t) := \int_0^t g(s)ds$ . Then, there exists a solution  $u \in V$  to the equation  $\mathcal{F}(u) = 0$  satisfying

$$\|u - \hat{u}\|_V \leq \alpha. \quad (10)$$

The solution is moreover unique under the side condition (10).

In the rest of this section, we consider the application of Theorem 2.1 to (4). Note that, in this case, the Fréchet derivative  $\mathcal{F}'_{\hat{u}}$  of  $\mathcal{F}$  at  $\hat{u} \in V$  is given by

$$\langle \mathcal{F}'_{\hat{u}} u, v \rangle = (\nabla u, \nabla v) - p \left( |\hat{u}|^{p-1} u, v \right) \quad \text{for } u, v \in V.$$

### Residual bound $\delta$

For  $\hat{u} \in V$  satisfying  $\Delta \hat{u} \in L^2(\Omega)$ , the residual bound  $\delta$  is computed as

$$C_2 \left\| \Delta \hat{u} + |\hat{u}|^{p-1} \hat{u} \right\|_{L^2(\Omega)};$$

the  $L^2$ -norm can be computed by a numerical integration method with verification (see Section 3 for details).

### Bound $K$ for the operator norm of $\mathcal{F}'_{\hat{u}}^{-1}$

In addition, we compute a bound  $K$  for the operator norm of  $\mathcal{F}'_{\hat{u}}^{-1}$  by the following theorem, proving simultaneously that this inverse operator exists and is defined on the whole of  $V^*$ .

**Theorem 2.2** ([17]). *Let  $\Phi : V \rightarrow V^*$  be the canonical isometric isomorphism, i.e.,  $\Phi$  is given by*

$$\langle \Phi u, v \rangle := (u, v)_V = (\nabla u, \nabla v) \quad \text{for } u, v \in V.$$

If

$$\mu_0 := \min \{ |\mu| : \mu \in \sigma_p(\Phi^{-1} \mathcal{F}'_{\hat{u}}) \cup \{1\} \} > 0, \quad (11)$$

with  $\sigma_p$  denoting the point spectrum, then the inverse of  $\mathcal{F}'_{\hat{u}}$  exists and

$$\|\mathcal{F}'_{\hat{u}}^{-1}\|_{B(V^*, V)} \leq \mu_0^{-1}. \quad (12)$$

*Proof.* We prove this theorem by adapting a theory of Fredholm operators, i.e., we have recourse to the fact that the injectivity and the surjectivity of a Fredholm operator are equivalent.

The operator  $N := \Phi - \mathcal{F}'_{\hat{u}}$  from  $V$  to  $V^*$  is given by  $\langle Nu, v \rangle = p(|\hat{u}|^{p-1}u, v)$  for all  $u, v \in V$ . Thus, actually  $N$  maps  $V$  into  $L^2(\Omega)$ ; note that  $p \leq 2$  and  $n \leq 3$ . Hence  $N : V \rightarrow V^*$  is compact, owing to the compactness of the embedding  $L^2(\Omega) \hookrightarrow V^*$ . Therefore,  $\mathcal{F}'_{\hat{u}}$  is a Fredholm operator, and the spectrum  $\sigma(\Phi^{-1}\mathcal{F}'_{\hat{u}})$  of  $\Phi^{-1}\mathcal{F}'_{\hat{u}}$  is given by

$$\sigma(\Phi^{-1}\mathcal{F}'_{\hat{u}}) = 1 - \sigma(\Phi^{-1}N) = 1 - \{\sigma_p(\Phi^{-1}N) \cup \{0\}\} = \sigma_p(\Phi^{-1}\mathcal{F}'_{\hat{u}}) \cup \{1\}.$$

Since  $\Phi^{-1}\mathcal{F}'_{\hat{u}}$  is self-adjoint, we have, for all  $u \in V$ ,

$$\|\mathcal{F}'_{\hat{u}}u\|_{V^*}^2 = \|\Phi^{-1}\mathcal{F}'_{\hat{u}}\|_V^2 = \int_{-\infty}^{\infty} \mu^2 d(E_\mu u, u)_V \geq \mu_0^2 \int_{-\infty}^{\infty} d(E_\mu u, u)_V = \mu_0^2 \|u\|_V^2,$$

where  $E_\mu$  is the resolution of the identity of  $\Phi^{-1}\mathcal{F}'_{\hat{u}}$ . Hence,  $\mathcal{F}'_{\hat{u}}$  is one to one, and therefore is also onto. This implies (12).  $\square$

**Remark 2.3.** *The property of the spectrum of  $\Phi^{-1}\mathcal{F}'_{\hat{u}}$  is more precisely discussed in [17, Section 3.3].*

The eigenvalue problem  $\Phi^{-1}\mathcal{F}'_{\hat{u}}u = \mu u$  in  $V$  is equivalent to

$$(\nabla u, \nabla v) - p(|\hat{u}|^{p-1}u, v) = \mu(\nabla u, \nabla v) \quad \text{for all } v \in V.$$

Since  $\mu = 1$  is already known to be in  $\sigma(\Phi^{-1}\mathcal{F}'_{\hat{u}})$ , it suffices to look for eigenvalues  $\mu \neq 1$ . By setting  $\lambda = (1 - \mu)^{-1}$ , we further transform this eigenvalue problem into

$$\text{Find } u \in V \text{ and } \lambda \in \mathbb{R} \text{ s.t. } (\nabla u, \nabla v) = \lambda(p|\hat{u}|^{p-1}u, v) \quad \text{for all } v \in V. \quad (13)$$

When we assume that  $\hat{u}(x) \neq 0$  for almost all  $x \in \Omega$  (which will be true by numerical construction in our example), i.e., we have  $|\hat{u}|^{p-1} > 0$  a.e. in  $\Omega$ , then (13) is a regular eigenvalue problem, the spectrum of which consists of a sequence  $\{\lambda_k\}_{k=1}^{\infty}$  of eigenvalues converging to  $+\infty$ . In order to compute  $K$  on the basis of Theorem 2.2, we concretely enclose the eigenvalue  $\lambda$  of (13) that minimizes the corresponding absolute value of  $|\mu| (= |1 - \lambda^{-1}|)$ , by considering the following approximate eigenvalue problem

$$\text{Find } u \in V_N \text{ and } \lambda^N \in \mathbb{R} \text{ s.t. } (\nabla u_N, \nabla v_N) = \lambda^N(p|\hat{u}|^{p-1}u_N, v_N) \quad \text{for all } v_N \in V_N, \quad (14)$$

where  $V_N$  is a finite-dimensional subspace of  $V$ . Note that (14) amounts to a matrix eigenvalue problem, the eigenvalues of which can easily be enclosed by verified numerical linear algebra (see, e.g., [2, 18, 8]).

To estimate the error between the  $k$ th eigenvalue  $\lambda_k$  of (13) and the  $k$ th eigenvalue  $\lambda_k^N$  of (14), we consider the weak formulation of the Poisson equation

$$(\nabla u, \nabla v) = (g, v) \quad \text{for all } v \in V \quad (15)$$

for given  $g \in L^2(\Omega)$ ; it is well known that this equation has a unique solution  $u \in V$  for each  $g \in L^2(\Omega)$ . Moreover, we introduce the orthogonal projection  $P_N : V \rightarrow V_N$  defined by

$$(P_N u - u, v_N)_V = 0 \quad \text{for all } u \in V \text{ and } v_N \in V_N.$$

The following theorem enables us to estimate the error between  $\lambda_k$  and  $\lambda_k^N$ .

**Theorem 2.4** ([24, 7]). Suppose that  $\hat{u} \in L^\infty(\Omega)$ , and let  $C_N$  denote a positive number such that

$$\|u_g - P_N u_g\|_V \leq C_N \|g\|_{L^2(\Omega)} \quad (16)$$

for any  $g \in L^2(\Omega)$  and the corresponding solution  $u_g \in V$  to (15). Then,

$$\frac{\lambda_k^N}{\lambda_k^N C_N^2 \|p|\hat{u}|^{p-1}\|_{L^\infty(\Omega)} + 1} \leq \lambda_k \leq \lambda_k^N.$$

The right inequality is well known as Rayleigh-Ritz bound, which is derived from the min-max principle:

$$\lambda_k = \min_{H_k \subset V} \left( \max_{v \in H_k \setminus \{0\}} \frac{\|\nabla v\|_{L^2(\Omega)}^2}{\|av\|_{L^2(\Omega)}^2} \right) \leq \lambda_k^N,$$

where we set  $a = \sqrt{p|\hat{u}|^{p-1}}$  and the minimum is taken over all  $k$ -dimensional subspaces  $H_k$  of  $V$ . Moreover, proofs of the left inequality can be found in [24, 7]. Assuming the  $H^2$ -regularity of solutions to (15) (e.g., when  $\Omega$  is convex [3, Section 3.3]), [24, Theorem 4] ensures the left inequality. A more general statement, that does not require the  $H^2$ -regularity, can be found in [7, Theorem 2.1].

**Remark 2.5.** When the  $H^2$ -regularity of solutions to (15) is confirmed a priori, e.g., when  $\Omega$  is convex [3, Section 3.3], (16) can be replaced by

$$\|u - P_N u\|_V \leq C_N \|-\Delta u\|_{L^2(\Omega)} \quad \text{for all } u \in H^2(\Omega) \cap V. \quad (17)$$

The computation of a concrete value of  $C_N$  for a given subspace  $V_N$  will be discussed in Section 5.

### Lipschitz bound $g$ for $\mathcal{F}'_{\hat{u}}$

Furthermore, a concrete construction of a function  $g$  satisfying (8) and (9) is important for our verification process. The following lemma is required for the construction.

**Lemma 2.6.** For  $a, b \in \mathbb{R}$  and  $q \in (0, 1)$ ,

$$||a + b|^q - |a|^q| \leq |b|^q.$$

*Proof.* For  $\alpha, \beta \in [0, \infty)$  we have

$$(\alpha + \beta)^q - \alpha^q = q \int_0^\beta (\alpha + t)^{q-1} dt \leq q \int_0^\beta t^{q-1} dt = \beta^q,$$

which readily gives

$$|a + b|^q - |a|^q \leq |b|^q \quad \text{for } a, b \in \mathbb{R}.$$

Redefining terms, this inequality also implies

$$|a|^q - |a + b|^q = |(a + b) + (-b)|^q - |a + b|^q \leq |-b|^q = |b|^q \quad \text{for } a, b \in \mathbb{R}$$

and hence the assertion.  $\square$

The following theorem gives us a concrete construction of the function  $g$  in Theorem 2.1 for the nonlinearity  $f(u) = |u|^{p-1}u$  ( $1 < p < 2$ ).

**Theorem 2.7.** *For  $1 < p < 2$ , we may select*

$$g(t) = pC_rC_sC_{q(p-1)}^{p-1}t^{p-1} \quad (18)$$

to satisfy (8) and (9) in Theorem 2.1, where  $q, r, s$  are positive numbers that satisfy  $q^{-1} + r^{-1} + s^{-1} = 1$  and  $q(p-1) \geq 1$ .

*Proof.* For fixed  $\hat{u} \in V$ , the left hand side of (8) is written as

$$\|\mathcal{F}'_{\hat{u}+u} - \mathcal{F}'_{\hat{u}}\|_{\mathcal{B}(V, V^*)} = p \sup_{v, \phi \in V \setminus \{0\}} \frac{\left| \left( (|\hat{u} + u|^{p-1} - |\hat{u}|^{p-1})v, \phi \right) \right|}{\|v\|_V \|\phi\|_V}.$$

Moreover, we have

$$\begin{aligned} \left| \left( (|\hat{u} + u|^{p-1} - |\hat{u}|^{p-1})v, \phi \right) \right| &\leq \left\| |\hat{u} + u|^{p-1} - |\hat{u}|^{p-1} \right\|_{L^q(\Omega)} \|v\|_{L^r(\Omega)} \|\phi\|_{L^s(\Omega)} \\ &\leq C_r C_s \left\| |\hat{u} + u|^{p-1} - |\hat{u}|^{p-1} \right\|_{L^q(\Omega)} \|v\|_V \|\phi\|_V, \end{aligned}$$

and, owing to Lemma 2.6,

$$\begin{aligned} \left\| |\hat{u} + u|^{p-1} - |\hat{u}|^{p-1} \right\|_{L^q(\Omega)} &= \left( \int_{\Omega} \left| |\hat{u}(x) + u(x)|^{p-1} - |\hat{u}(x)|^{p-1} \right|^q dx \right)^{1/q} \\ &\leq \left( \int_{\Omega} |u(x)|^{q(p-1)} dx \right)^{1/q} = \|u\|_{L^{q(p-1)}(\Omega)}^{p-1}. \end{aligned}$$

Therefore, it follows that

$$\|\mathcal{F}'_{\hat{u}+u} - \mathcal{F}'_{\hat{u}}\|_{\mathcal{B}(V, V^*)} \leq pC_rC_sC_{q(p-1)}^{p-1} \|u\|_V^{p-1} = g(\|u\|_V).$$

□

## 2.2 $L^\infty$ error estimation

In this subsection, we discuss a method that gives an  $L^\infty$  error bound for a solution to (4) from a known  $H_0^1$  error bound, that is, we compute a concrete bound for  $\|u - \hat{u}\|_{L^\infty(\Omega)}$  for a solution  $u \in V$  to (4) satisfying

$$\|u - \hat{u}\|_V \leq \varepsilon \quad (19)$$

with  $\varepsilon > 0$  and  $\hat{u} \in V$ . To obtain such an error estimation, we assume that  $\Omega$  is convex and polygonal; this condition gives the  $H^2$ -regularity of solutions to (4) (and therefore, ensures their boundedness) a priori. To be precise, when  $\Omega$  is a convex polygonal domain, a weak solution  $u \in V$  to (15) with  $g \in L^2(\Omega)$  is  $H^2$ -regular (see, e.g., [3, Section 3.3]). A solution  $u$  satisfying (19) can be written in the form  $u = \hat{u} + \varepsilon\omega$  with some  $\omega \in V$ ,  $\|\omega\|_V \leq 1$ . Moreover,  $\omega$  satisfies

$$\begin{cases} -\Delta \varepsilon \omega = |\hat{u} + \varepsilon \omega|^{p-1} (\hat{u} + \varepsilon \omega) + \Delta \hat{u} & \text{in } \Omega, \\ \omega = 0 & \text{on } \partial\Omega, \end{cases}$$

and therefore is also  $H^2$ -regular if  $\Delta \hat{u} \in L^2(\Omega)$ . We then use the following theorem to obtain an  $L^\infty$  error estimation.

**Theorem 2.8** ([14]). For all  $u \in H^2(\Omega)$ ,

$$\|u\|_{L^\infty(\Omega)} \leq c_0 \|u\|_{L^2(\Omega)} + c_1 \|\nabla u\|_{L^2(\Omega)} + c_2 \|u_{xx}\|_{L^2(\Omega)}$$

with

$$c_j = \frac{\gamma_j}{|\overline{\Omega}|} \left[ \max_{x_0 \in \overline{\Omega}} \int_{\overline{\Omega}} |x - x_0|^{2j} dx \right]^{1/2}, \quad (j = 0, 1, 2),$$

where  $u_{xx}$  denotes the Hesse matrix of  $u$ ,  $|\overline{\Omega}|$  is the measure of  $\overline{\Omega}$ , and

$$\gamma_0 = 1, \quad \gamma_1 = 1.1548, \quad \gamma_2 = 0.22361 \quad \text{if } n = 2.$$

For  $n = 3$ , other values of  $\gamma_0$ ,  $\gamma_1$ , and  $\gamma_2$  have to be chosen (see [14]).

**Remark 2.9.** The norm of the Hesse matrix of  $u$  is precisely defined by

$$\|u_{xx}\|_{L^2(\Omega)} = \sqrt{\sum_{i,j=1}^2 \left\| \frac{\partial^2 u}{\partial x_i \partial x_j} \right\|_{L^2(\Omega)}^2}.$$

Moreover, since  $\Omega$  is polygonal,  $\|u_{xx}\|_{L^2(\Omega)} = \|\Delta u\|_{L^2(\Omega)}$  for all  $u \in H^2(\Omega) \cap V$  (see, e.g., [3]).

**Remark 2.10.** Concrete values of each  $c_j$  are provided for some special domains  $\Omega$  in [14, 15]. According to these papers, one can choose, for  $\Omega = (0, 1)^2$ ,

$$c_0 = \gamma_0, \quad c_1 = \sqrt{\frac{2}{3}} \gamma_1, \quad \text{and} \quad c_2 = \frac{\gamma_3}{3} \sqrt{\frac{28}{5}}.$$

Applying Theorem 2.8, we obtain the following corollary.

**Corollary 2.11.** Let  $u$  be a solution to (4) satisfying (19) with  $\hat{u} \in V$  such that  $\Delta \hat{u} \in L^2(\Omega)$ . Moreover, let  $c_0$ ,  $c_1$ , and  $c_2$  be as in Theorem 2.8, and  $p' := 2(p-1)$ . Then,

$$\begin{aligned} & \|u - \hat{u}\|_{L^\infty(\Omega)} \\ & \leq c_0 C_2 \varepsilon + c_1 \varepsilon + c_2 \left\{ \max\{1, 2^{\frac{p'-1}{2}}\} p \varepsilon C_q \sqrt{\|\hat{u}\|_{L^{rp'}(\Omega)}^{p'} + \frac{\varepsilon^{p'}}{p'+1} C_{rp'}^{p'} + \|\Delta \hat{u} + |\hat{u}|^{p-1} \hat{u}\|_{L^2(\Omega)}} \right\} \end{aligned} \quad (20)$$

holds for any  $q$  and  $r$  satisfying  $q \geq 2$ ,  $r \geq (p-1)^{-1}$ , and  $2q^{-1} + r^{-1} = 1$ .

*Proof.* Due to Theorem 2.8, we have

$$\begin{aligned} \|u - \hat{u}\|_{L^\infty(\Omega)} &= \varepsilon \|\omega\|_{L^\infty(\Omega)} \\ &\leq \varepsilon \left( c_0 \|\omega\|_{L^2(\Omega)} + c_1 \|\omega\|_V + c_2 \|\Delta \omega\|_{L^2(\Omega)} \right) \\ &\leq \varepsilon \left( c_0 C_2 + c_1 + c_2 \|\Delta \omega\|_{L^2(\Omega)} \right). \end{aligned}$$

The last term  $\|\Delta \omega\|_{L^2(\Omega)}$  is estimated by

$$\varepsilon \|\Delta \omega\|_{L^2(\Omega)} = \left\| |\hat{u} + \varepsilon \omega|^{p-1} (\hat{u} + \varepsilon \omega) + \Delta \hat{u} \right\|_{L^2(\Omega)}$$

$$\begin{aligned}
&= \left\| |\hat{u} + \varepsilon\omega|^{p-1} (\hat{u} + \varepsilon\omega) - |\hat{u}|^{p-1} \hat{u} + |\hat{u}|^{p-1} \hat{u} + \Delta\hat{u} \right\|_{L^2(\Omega)} \\
&\leq \left\| |\hat{u} + \varepsilon\omega|^{p-1} (\hat{u} + \varepsilon\omega) - |\hat{u}|^{p-1} \hat{u} \right\|_{L^2(\Omega)} + \left\| \Delta\hat{u} + |\hat{u}|^{p-1} \hat{u} \right\|_{L^2(\Omega)}
\end{aligned}$$

Since the mean value theorem ensures that

$$\begin{aligned}
&\int_{\Omega} \left( |\hat{u}(x) + \varepsilon\omega(x)|^{p-1} (\hat{u}(x) + \varepsilon\omega(x)) - |\hat{u}(x)|^{p-1} \hat{u}(x) \right)^2 dx \\
&= \int_{\Omega} \left( \varepsilon p \omega(x) \int_0^1 |\hat{u}(x) + \varepsilon t \omega(x)|^{p-1} dt \right)^2 dx \\
&\leq p^2 \varepsilon^2 \int_{\Omega} \omega(x)^2 \int_0^1 |\hat{u}(x) + \varepsilon t \omega(x)|^{p'} dt dx \\
&= p^2 \varepsilon^2 \int_0^1 \int_{\Omega} \omega(x)^2 |\hat{u}(x) + \varepsilon t \omega(x)|^{p'} dx dt \\
&\leq p^2 \varepsilon^2 \|\omega\|_{L^q(\Omega)}^2 \int_0^1 \left\| |\hat{u} + \varepsilon \omega t|^{p'} \right\|_{L^{r'}(\Omega)} dt \\
&= p^2 \varepsilon^2 \|\omega\|_{L^q(\Omega)}^2 \int_0^1 \|\hat{u} + \varepsilon \omega t\|_{L^{rp'}(\Omega)}^{p'} dt \\
&\leq p^2 \varepsilon^2 \|\omega\|_{L^q(\Omega)}^2 \int_0^1 \left( \|\hat{u}\|_{L^{rp'}(\Omega)} + t \varepsilon \|\omega\|_{L^{rp'}(\Omega)} \right)^{p'} dt \\
&\leq \max\{1, 2^{p'-1}\} p^2 \varepsilon^2 \|\omega\|_{L^q(\Omega)}^2 \left\{ \|\hat{u}\|_{L^{rp'}(\Omega)}^{p'} + \int_0^1 \left( t \varepsilon \|\omega\|_{L^{rp'}(\Omega)} \right)^{p'} dt \right\} \\
&= \max\{1, 2^{p'-1}\} p^2 \varepsilon^2 \|\omega\|_{L^q(\Omega)}^2 \left( \|\hat{u}\|_{L^{rp'}(\Omega)}^{p'} + \frac{\varepsilon^{p'}}{p' + 1} \|\omega\|_{L^{rp'}(\Omega)}^{p'} \right) \\
&\leq \max\{1, 2^{p'-1}\} p^2 \varepsilon^2 C_q^2 \left( \|\hat{u}\|_{L^{rp'}(\Omega)}^{p'} + \frac{\varepsilon^{p'}}{p' + 1} C_{rp'}^{p'} \right),
\end{aligned}$$

it follows that

$$\varepsilon \|\Delta\omega\|_{L^2(\Omega)} \leq \max\{1, 2^{\frac{p'-1}{2}}\} p \varepsilon C_q \sqrt{\|\hat{u}\|_{L^{rp'}(\Omega)}^{p'} + \frac{\varepsilon^{p'}}{p' + 1} C_{rp'}^{p'}} + \left\| \Delta\hat{u} + |\hat{u}|^{p-1} \hat{u} \right\|_{L^2(\Omega)}.$$

Consequently, the  $L^\infty$  error of  $u$  is estimated as asserted in (20).  $\square$

### 3 Verified numerical integration

To apply Theorem 2.1 to problem (4), one has to construct a “good” approximation  $\hat{u} \in V$  of a solution to (4) such that  $\delta$  in (6) is sufficiently small. In this paper, we assume that such an approximation  $\hat{u}$  is constructed by a finite linear combination of basis functions  $\{\phi_i\}_{i=1}^\infty$  that span  $V$ , where each  $\phi_i$  is in  $C^\infty(\overline{\Omega})$  (and therefore,  $\hat{u} \in C^\infty(\overline{\Omega})$ ). To obtain concrete bounds for  $\delta$  and  $K$  required in Theorem 2.1, one has to compute, in particular,  $(\Delta\hat{u}, |\hat{u}|^{p-1} \hat{u})_{L^2(\Omega)}$  and  $(\phi_i, |\hat{u}|^{p-1} \phi_j)_{L^2(\Omega)}$  with verification (recall that  $1 < p < 2$  which makes this integration non-trivial).

In this section, for the square  $\Omega_s = (0, 1)^2 \subset \mathbb{R}^2$ ,  $0 < q < 1$ , and  $\eta, \xi \in C^\infty(\overline{\Omega}_s)$ , we propose a method for computing the integral

$$I = \int_{\Omega_s} \{\eta(x, y)\}^q \xi(x, y) dx dy,$$



in verified form, i.e., for computing an enclosure for this integral, where we assume that  $\eta > 0$  in  $\Omega_s$  and  $\eta = 0$  on  $\partial\Omega_s$ ; indeed, we later select  $\eta = \hat{u}$ , an approximate solution to (3), which has these properties. We prove the positivity of  $\hat{u}$  in  $\Omega$  using the procedures described in Subsections 3.1 to 3.3.

There are some verified integration methods that can be applied to such an integration, under the assumption that  $\eta > 0$  on the whole closure of a domain  $\Omega \subset \mathbb{R}^2$  (see, e.g, [19]). However, since here the derivative of  $\{\eta(\cdot, \cdot)\}^q : \Omega \rightarrow \mathbb{R}$  is in general not bounded near the boundary  $\partial\Omega$ , where  $\eta$  vanishes, previous methods cannot be applied in our situation. To overcome this difficulty, we employ a Taylor expansion based method as follows:

We first divide  $\Omega_s$  into four sub-squares, and consider the integration over  $\Omega_{s/4} := (0, 1/2)^2$ ; integration over the three other parts can be carried out similarly, after translation and rotation such that  $\eta = 0$  on both the left and the lower edge. Moreover, we divide  $\overline{\Omega_{s/4}}$  into closed rectangles that are grouped into four types ( $S_{1,1}$ ,  $S_{1,0}$ ,  $S_{0,1}$ , and  $S_{0,0}$ ) as in Fig. 1. These types

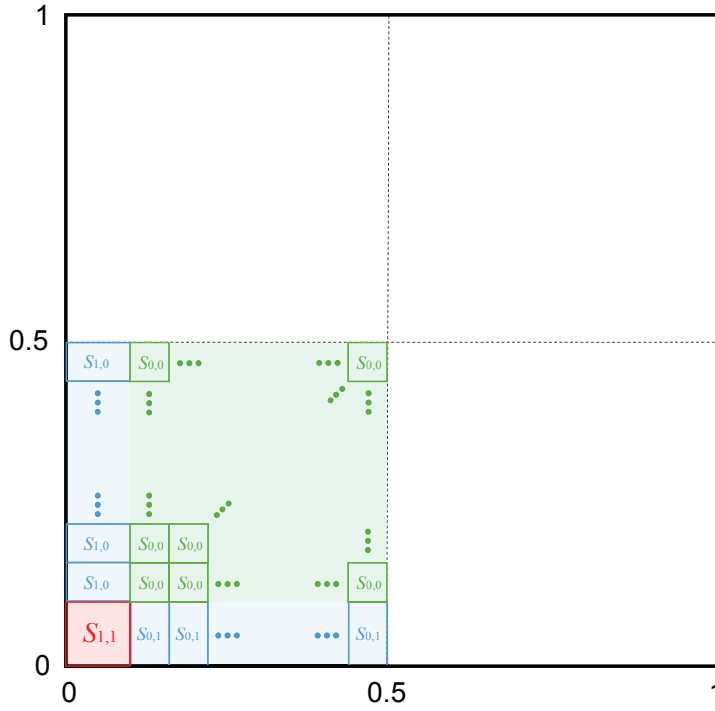


Figure 1: Division of the domain.

of rectangles have the following properties:

$S_{1,1}$ : a square where  $\eta$  is zero on both the left and the lower edge;

$S_{0,1}$ : rectangles where  $\eta$  is zero only on the lower edge;

$S_{1,0}$ : rectangles where  $\eta$  is zero only on the left edge;

$S_{0,0}$ : squares where  $\eta > 0$ .

Then, the integration over  $\Omega_{s/4}$  can be expressed by summation of integrations over the above four types of rectangles. Therefore, we discuss an integration method over the four types of domains. Hereafter, we employ the notation  $\Lambda_n^{1,1} = \{(i, j) \in \mathbb{N}^2 : i \leq n, j \leq n\}$ ,  $\Lambda_n^{0,1} = \{(i, j) \in \mathbb{N}_0 \times \mathbb{N} : i \leq n, j \leq n\}$ ,  $\Lambda_n^{1,0} = \{(i, j) \in \mathbb{N} \times \mathbb{N}_0 : i \leq n, j \leq n\}$ , and  $\Lambda_n = \{(i, j) \in \mathbb{N}_0^2 : i \leq n, j \leq n\}$ , where  $\mathbb{N} = \{1, 2, 3 \dots\}$  and  $\mathbb{N}_0 = \{0, 1, 2 \dots\}$ .

### 3.1 Integration over $S_{1,1}$

On the basis of the Taylor expansion around the lower left corner  $(0,0)$ , we enclose  $\eta(x,y)$  as

$$\eta(x,y) \in \sum_{(i,j) \in \Lambda_{n-1}^{1,1}} a_{i,j} x^i y^j + \sum_{(i,j) \in \Lambda_n^{1,1} \setminus \Lambda_{n-1}^{1,1}} [\underline{a}_{i,j}, \bar{a}_{i,j}] x^i y^j, \quad (21)$$

for  $(x,y) \in S_{1,1}$ , where  $a_{i,j}, \underline{a}_{i,j}, \bar{a}_{i,j} \in \mathbb{R}$ ,  $\underline{a}_{i,j} \leq \bar{a}_{i,j}$ . In Section B, we will introduce a numerical method (Type-II PSA) for deriving such an enclosure. We then denote

$$[\eta_{1,1}(x,y)] := \sum_{(i,j) \in \Lambda_{n-1}^{1,1}} a_{i,j} x^{i-1} y^{j-1} + \sum_{(i,j) \in \Lambda_n^{1,1} \setminus \Lambda_{n-1}^{1,1}} [\underline{a}_{i,j}, \bar{a}_{i,j}] x^{i-1} y^{j-1},$$

which more precisely means the set of all continuous functions  $w$  over  $S_{1,1}$  such that  $w(x,y) \in [\eta_{1,1}(x,y)]$  for all  $(x,y) \in S_{1,1}$ . Therefore  $\eta(x,y) \in xy[\eta_{1,1}(x,y)]$ .

We moreover assume that  $[\eta_{1,1}(x,y)]$  is positive in  $S_{1,1}$  (i.e.,  $z > 0$  for all  $z \in [\eta_{1,1}(x,y)]$ ,  $(x,y) \in S_{1,1}$ ); if  $S_{1,1}$  is sufficiently small and  $n$  is sufficiently large, this positivity condition is expected to hold for  $\eta = \hat{u}$  (in the actual computation, this condition will be numerically checked by suitable interval arithmetic techniques [9, 18]). Under this assumption, we use Type-II PSA first to enclose  $[\eta_{1,1}(x,y)]^q$ , and then, in a second step, to enclose  $[\eta_{1,1}(x,y)]^q \xi(x,y)$  as

$$[\eta_{1,1}(x,y)]^q \xi(x,y) \in \sum_{(i,j) \in \Lambda_{n-1}} b_{i,j} x^i y^j + \sum_{(i,j) \in \Lambda_n \setminus \Lambda_{n-1}} [\underline{b}_{i,j}, \bar{b}_{i,j}] x^i y^j,$$

where  $b_{i,j}, \underline{b}_{i,j}, \bar{b}_{i,j} \in \mathbb{R}$ ,  $\underline{b}_{i,j} \leq \bar{b}_{i,j}$ . Hence, the integration over  $S_{1,1}$  is enclosed as

$$\begin{aligned} & \int_{S_{1,1}} \{\eta(x,y)\}^q \xi(x,y) dx dy \\ & \in \sum_{(i,j) \in \Lambda_{n-1}} \int_{S_{1,1}} b_{i,j} x^{i+q} y^{j+q} dx dy + \sum_{(i,j) \in \Lambda_n \setminus \Lambda_{n-1}} \int_{S_{1,1}} [\underline{b}_{i,j}, \bar{b}_{i,j}] x^{i+q} y^{j+q} dx dy. \end{aligned} \quad (22)$$

**Remark 3.1.** We wish to make a remark about integration of a set of continuous functions as in (22), i.e., we provide an explanation of verified integration of

$$\int_{y_1}^{y_2} \int_{x_1}^{x_2} a x^p y^q dx dy,$$

where generally,  $a, p, q, x_1, x_2, y_1$ , and  $y_2$  are real intervals. Note that  $a x^p y^q$  precisely means the set of all continuous functions  $w$  over  $(\underline{x_1}, \overline{x_2}) \times (\underline{y_1}, \overline{y_2})$  such that  $w(x,y) \in a x^p y^q$  for all  $(x,y) \in (\underline{x_1}, \overline{x_2}) \times (\underline{y_1}, \overline{y_2})$ , where we denote  $\underline{z} = \inf z$  and  $\overline{z} = \sup z$  for an interval  $z$ . While formally the integral is simply computed as

$$\begin{aligned} \int_{y_1}^{y_2} \int_{x_1}^{x_2} a x^p y^q dx dy &= \frac{a}{(p+1)(q+1)} y_2^{q+1} x_2^{p+1} - \frac{a}{(p+1)(q+1)} y_2^{q+1} x_1^{p+1} \\ &\quad - \left( \frac{a}{(p+1)(q+1)} y_1^{q+1} x_2^{p+1} - \frac{a}{(p+1)(q+1)} y_1^{q+1} x_1^{p+1} \right), \end{aligned}$$

one has to compute the above formula in correct order using suitable interval arithmetic techniques, because the distributive law does not hold in interval arithmetics. For example,  $\int_{-1}^1 [0.8, 1] dx$  is not zero, but is correctly computed as

$$\int_{-1}^1 [0.8, 1] dx = [[0.4, 0.5] x^2]_{-1}^1 = [0.4, 0.5] - [0.4, 0.5] = [-0.1, 0.1].$$

### 3.2 Integration over $S_{0,1}$ and $S_{1,0}$

Let  $(x_0, 0)$  be the midpoint of the lower edge of  $S_{0,1}$ . We denote  $\eta^*(x, y) := \eta(x + x_0, y)$ ,  $\xi^*(x, y) := \xi(x + x_0, y)$ , and  $S_{0,1}^* := S_{0,1} - (x_0, 0)$ . Since we have

$$\int_{S_{0,1}} \{\eta(x, y)\}^q \xi(x, y) dx dy = \int_{S_{0,1}^*} \{\eta^*(x, y)\}^q \xi^*(x, y) dx dy,$$

we consider the right integral in this subsection.

By Taylor expanding  $\eta^*(x, y)$  around the midpoint  $(0, 0)$  of the lower edge of  $S_{0,1}^*$ , we enclose  $\eta^*(x, y)$  as

$$\eta^*(x, y) \in \sum_{(i,j) \in \Lambda_{n-1}^{0,1}} a_{i,j} x^i y^j + \sum_{(i,j) \in \Lambda_n^{0,1} \setminus \Lambda_{n-1}^{0,1}} [\underline{a}_{i,j}, \bar{a}_{i,j}] x^i y^j,$$

for  $(x, y) \in S_{0,1}^*$ , where  $a_{i,j}, \underline{a}_{i,j}, \bar{a}_{i,j} \in \mathbb{R}$ ,  $\underline{a}_{i,j} \leq \bar{a}_{i,j}$ . We then denote

$$[\eta_{0,1}^*(x, y)] := \sum_{(i,j) \in \Lambda_{n-1}^{0,1}} a_{i,j} x^i y^{j-1} + \sum_{(i,j) \in \Lambda_n^{0,1} \setminus \Lambda_{n-1}^{0,1}} [\underline{a}_{i,j}, \bar{a}_{i,j}] x^i y^{j-1}$$

(therefore,  $\eta^*(x, y) \in y[\eta_{0,1}^*(x, y)]$ ), and again assume that  $[\eta_{0,1}^*(x, y)]$  is positive in  $S_{0,1}^*$ . We then again enclose  $[\eta_{0,1}^*(x, y)]^q \xi^*(x, y)$  as

$$[\eta_{0,1}^*(x, y)]^q \xi^*(x, y) \in \sum_{(i,j) \in \Lambda_{n-1}} b_{i,j} x^i y^j + \sum_{(i,j) \in \Lambda_n \setminus \Lambda_{n-1}} [\underline{b}_{i,j}, \bar{b}_{i,j}] x^i y^j,$$

where  $b_{i,j}, \underline{b}_{i,j}, \bar{b}_{i,j} \in \mathbb{R}$ ,  $\underline{b}_{i,j} \leq \bar{b}_{i,j}$ . Thus, we can enclose the integral over  $S_{0,1}^*$  as

$$\begin{aligned} & \int_{S_{0,1}^*} \{\eta^*(x, y)\}^q \xi^*(x, y) dx dy \\ & \in \sum_{(i,j) \in \Lambda_{n-1}} \int_{S_{0,1}^*} b_{i,j} x^i y^{j+q} dx dy + \sum_{(i,j) \in \Lambda_n \setminus \Lambda_{n-1}} \int_{S_{0,1}^*} [\underline{b}_{i,j}, \bar{b}_{i,j}] x^i y^{j+q} dx dy. \end{aligned}$$

The integration over  $S_{1,0}$  is carried out similarly by exchanging the roles of the variables  $x$  and  $y$ .

### 3.3 Integration over $S_{0,0}$

Let  $(x_0, y_0)$  be the center of  $S_{0,0}$ , and we re-define  $\eta^*(x, y) := \eta(x + x_0, y + y_0)$ ,  $\xi^*(x, y) := \xi(x + x_0, y + y_0)$ , and  $S_{0,0}^* := S_{0,0} - (x_0, y_0)$ . Since we have

$$\int_{S_{0,0}} \{\eta(x, y)\}^q \xi(x, y) dx dy = \int_{S_{0,0}^*} \{\eta^*(x, y)\}^q \xi^*(x, y) dx dy,$$

we consider the right integral in this subsection.

By Taylor expanding  $\eta^*(x, y)$  around the center  $(0, 0)$  of  $S_{0,0}^*$ , we have

$$\eta^*(x, y) \in [\eta_{0,0}^*(x, y)] := \sum_{(i,j) \in \Lambda_{n-1}} a_{i,j} x^i y^j + \sum_{(i,j) \in \Lambda_n \setminus \Lambda_{n-1}} [\underline{a}_{i,j}, \bar{a}_{i,j}] x^i y^j,$$

for  $(x, y) \in S_{0,0}^*$ , where  $a_{i,j}, \underline{a}_{i,j}, \bar{a}_{i,j} \in \mathbb{R}$ ,  $\underline{a}_{i,j} \leq \bar{a}_{i,j}$ . Assuming that  $[\eta_{0,0}^*(x, y)]$  is positive on  $S_{0,0}^*$  (since  $\eta > 0$  on  $S_{0,0}^*$ , this property is expected to hold if  $n$  is sufficiently large), we have

$$[\eta_{0,0}^*(x, y)]^q \xi^*(x, y) \in \sum_{(i,j) \in \Lambda_{n-1}} b_{i,j} x^i y^j + \sum_{(i,j) \in \Lambda_n \setminus \Lambda_{n-1}} [\underline{b}_{i,j}, \bar{b}_{i,j}] x^i y^j,$$

where  $b_{i,j}, \underline{b}_{i,j}, \bar{b}_{i,j} \in \mathbb{R}$ ,  $\underline{b}_{i,j} \leq \bar{b}_{i,j}$ . Thus, we have

$$\begin{aligned} & \int_{S_{0,0}^*} \{\eta^*(x, y)\}^q \xi^*(x, y) dx dy \\ & \in \sum_{(i,j) \in \Lambda_{n-1}} \int_{S_{0,0}^*} b_{i,j} x^i y^j dx dy + \sum_{(i,j) \in \Lambda_n \setminus \Lambda_{n-1}} \int_{S_{0,0}^*} [\underline{b}_{i,j}, \bar{b}_{i,j}] x^i y^j dx dy. \end{aligned}$$

**Remark 3.2.** *Integration over  $S_{0,0}$  can also be carried out using common methods (see, e.g., [19]).*

## 4 Verification of positiveness

One can prove positiveness of a (strong) solution to (4) using the following theorem.

**Theorem 4.1.** *Let  $\Omega$  be a bounded domain in  $\mathbb{R}^n$  ( $n = 1, 2, 3, \dots$ ). If a solution  $u \in C^2(\Omega) \cap C(\bar{\Omega})$  to (4) is positive in a nonempty subdomain  $\Omega' \subset \Omega$  and  $\sup\{(u_-(x))^{p-1} \mid x \in \Omega\} < \lambda_1(\Omega)$ , then  $u > 0$  in the original domain  $\Omega$ ; that is,  $u$  is also a solution to (3). Here,  $\lambda_1(\Omega) > 0$  is the first eigenvalue of the problem*

$$(\nabla u, \nabla v)_{L^2(\Omega)} = \lambda(u, v)_{L^2(\Omega)}, \quad \forall v \in V.$$

and  $u_-$  is defined by

$$u_-(x) := \begin{cases} -u(x), & u(x) < 0, \\ 0, & u(x) \geq 0. \end{cases}$$

A proof can be found in [22, 23].

## 5 Numerical example

In this section, we present a numerical example where a solution to (3) is numerically verified. All computations were carried out on a computer with Intel Xeon E7-4830 at 2.20 GHz  $\times$  40, 2 TB RAM, CentOS 6.6, and MATLAB 2012b. All rounding errors were strictly estimated using toolboxes—the INTLAB version 9 [18] and KV library version 0.4.16 [6]—for verified numerical computations. Therefore, the correctness of all results was mathematically guaranteed.

We consider the case in which  $p = 3/2$  and  $\Omega = \Omega_s := (0, 1)^2$ . By selecting  $q = r = 4$  and  $s = 2$  in Theorem 2.7, we may select

$$g(t) = \frac{3}{2} C_2^{\frac{3}{2}} C_4 t^{\frac{1}{2}} \quad (23)$$

to satisfy (8) and (9) in Theorem 2.1. Moreover, by selecting  $q = 4$  and  $r = 2$  in Corollary 2.11, we have

$$\|u - \hat{u}\|_{L^\infty(\Omega)} \leq c_0 C_2 \varepsilon + c_1 \varepsilon + c_2 \left\{ \frac{3}{2} \varepsilon C_4 \sqrt{\|\hat{u}\|_{L^2(\Omega)} + \frac{\varepsilon}{2} C_2} + \left\| \Delta \hat{u} + |\hat{u}|^{\frac{1}{2}} \hat{u} \right\|_{L^2(\Omega)} \right\}.$$

On the square  $\Omega_s$ , it is well known that  $C_2 = (\sqrt{2}\pi)^{-1}$ ; moreover, using Corollary A.2, we computed  $C_4 \leq 0.318309887$ .

We select a finite-dimensional subspace  $V_N$  of  $V$  as

$$V_N := \left\{ \sum_{(i,j) \in \Lambda_N^{1,1}} a_{i,j} \varphi_{i,j} : a_{i,j} \in \mathbb{R} \right\},$$

where  $\varphi_{i,j}(x, y) = \sin(i\pi x) \sin(j\pi y)$ . For this  $V_N$ , we may select  $C_N = (N+1)^{-1}\pi^{-1}$  to satisfy (16), because

$$\begin{aligned} \|u - P_N u\|_V^2 &= \|(u - P_N u)_x\|_{L^2(\Omega)}^2 + \|(u - P_N u)_y\|_{L^2(\Omega)}^2 \\ &= \sum_{(n,m) \in \Lambda_\infty^{1,1} \setminus \Lambda_N^{1,1}} a_{m,n}^2 (m^2 \pi^2 + n^2 \pi^2) \|\varphi_{m,n}\|_{L^2(\Omega)}^2 \\ &\leq \sum_{(n,m) \in \Lambda_\infty^{1,1} \setminus \Lambda_N^{1,1}} a_{m,n}^2 \frac{(m^2 \pi^2 + n^2 \pi^2)^2}{(N+1)^2 \pi^2} \|\varphi_{m,n}\|_{L^2(\Omega)}^2 \\ &\leq \frac{1}{(N+1)^2 \pi^2} \|-\Delta u\|_{L^2(\Omega)}^2. \end{aligned}$$

We are interested in finding a reflection symmetric solution, and hence restricted the solution space to the following subspace of  $V$ :

$$\left\{ u \in V : u \text{ is symmetric with respect to } x = \frac{1}{2} \text{ and } y = \frac{1}{2} \right\}$$

endowed with the same topology as  $V$ . This restriction helped us to somewhat reduce the calculation quantity. Moreover, since eigenfunctions of (13) are now also restricted to symmetric functions, eigenvalues associated with anti-symmetric eigenfunction drop out of the minimization in (11), and so the constant  $K$  is possibly reduced. The other constants required in the process of the verification (i.e.,  $C_p$ ,  $\delta$ ) are not affected by the restriction.

We computed an approximate solution  $\hat{u}$  to (3), which is displayed in Fig. 2, with the Fourier-Galerkin method, i.e.,  $\hat{u}$  was put up in the form

$$\hat{u}(x, y) = \sum_{\substack{1 \leq i, j \leq N_u \\ i, j \text{ are odd}}} a_{i,j} \varphi_{i,j}(x, y),$$

where  $N_u = 60$ .

Using Theorem 2.1 and Corollary 2.11, we proved the existence of a solution  $u$  to (4) in an  $H_0^1$ -ball  $\overline{B}(\hat{u}, r_1; \|\cdot\|_V)$  and an  $L^\infty$ -ball  $\overline{B}(\hat{u}, r_2; \|\cdot\|_{L^\infty(\Omega)})$ , where  $\overline{B}(x, r; \|\cdot\|)$  denotes the closed ball whose center is  $x$ , and whose radius is  $r \geq 0$  with respect to the norm  $\|\cdot\|$ . Table 1 presents the verification result, which ensures positiveness of the verified solution  $u$  owing to the condition  $\sup\{\sqrt{u_-(x)} \mid x \in \Omega_s\} \leq \lambda_1 (= 2\pi^2)$ , and therefore, it is also a (strong) solution to (3). Here, the upper bound of  $\sup\{\sqrt{u_-(x)} \mid x \in \Omega_s\}$  was calculated by  $[\min\{\hat{u}(x) : x \in \Omega_s\} + r_2]^{p-1}$  with verification. Note also that  $u \in C^2(\Omega)$  by local regularity, and  $u \in C(\overline{\Omega})$  due to the embedding  $H^2(\Omega) \hookrightarrow C(\overline{\Omega})$ , which indeed allows application of Theorem 4.1.

## Appendix A Simple bounds for the needed embedding constants

The following theorem provides the best constant in the classical Sobolev inequality with critical exponents.

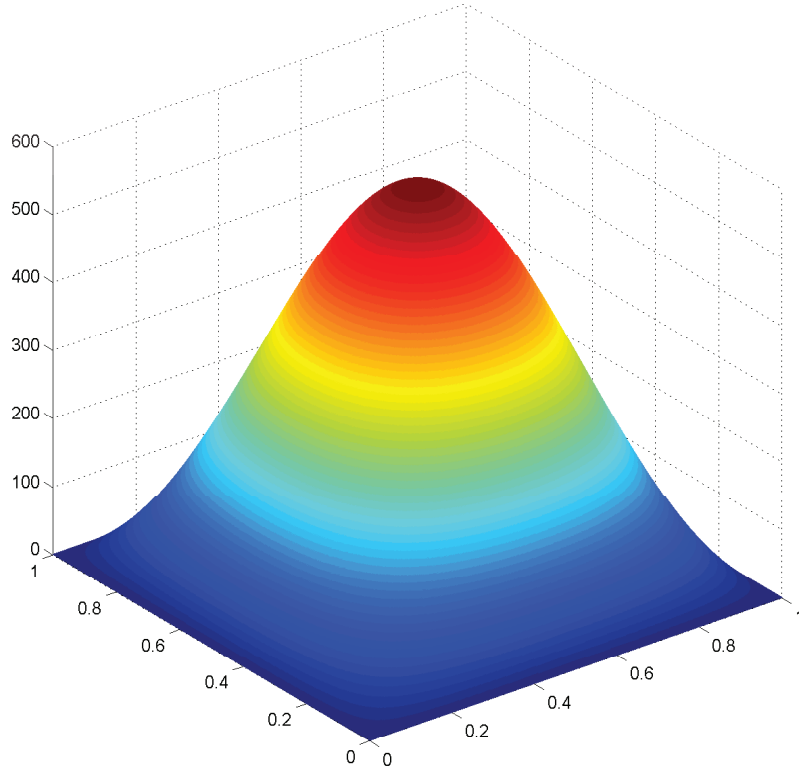


Figure 2: Approximate solution to (3) on  $\Omega_s := (0, 1)^2$ , the amplitude of which is proved to be in the interval  $[575.15, 575.61]$ .

Table 1: Verification result for (3).

$\ \Delta \hat{u} +  \hat{u} ^{\frac{1}{2}} \hat{u}\ _{L^2}$	$\delta$	$K$ ( $N = 14$ )	$r_1$	$r_2$	$\sup \sqrt{u_-(x)}$
$[0.8311281, 0.8314938]$	0.1871519	2.0000005	0.3909193	1.1462326	1.0706226

$\delta$ ,  $K$ : the constants required in Theorem 2.1.

The value of  $\|\Delta \hat{u} + |\hat{u}|^{\frac{1}{2}} \hat{u}\|_{L^2}$  is proved to be in the displayed interval. The other numerical values represent upper bounds of the corresponding constants.

**Theorem A.1** (T. Aubin [1] and G. Talenti [21]). *Let  $u$  be any function in  $W^{1,q}(\mathbb{R}^n)$  ( $n \geq 2$ ), where  $q$  is any real number such that  $1 < q < n$ . Moreover, set  $p = nq/(n - q)$ . Then,  $u \in L^p(\mathbb{R}^n)$  and*

$$\left( \int_{\mathbb{R}^n} |u(x)|^p dx \right)^{\frac{1}{p}} \leq T_p \left( \int_{\mathbb{R}^n} |\nabla u(x)|_2^q dx \right)^{\frac{1}{q}}$$

holds for

$$T_p = \pi^{-\frac{1}{2}} n^{-\frac{1}{q}} \left( \frac{q-1}{n-q} \right)^{1-\frac{1}{q}} \left\{ \frac{\Gamma(1 + \frac{n}{2}) \Gamma(n)}{\Gamma(\frac{n}{q}) \Gamma(1 + n - \frac{n}{q})} \right\}^{\frac{1}{n}}, \quad (24)$$

where  $|\nabla u|_2 = ((\partial u / \partial x_1)^2 + (\partial u / \partial x_2)^2 + \cdots + (\partial u / \partial x_n)^2)^{1/2}$ , and  $\Gamma$  denotes the gamma function.

The following corollary, obtained from Theorem A.1, provides a simple bound for the embedding constant from  $H_0^1(\Omega)$  to  $L^p(\Omega)$  for a bounded domain  $\Omega$ .

**Corollary A.2.** *Let  $\Omega \subset \mathbb{R}^n$  ( $n \geq 2$ ) be a bounded domain. Let  $p$  be a real number such that  $p \in (n/(n-1), 2n/(n-2)]$  if  $n \geq 3$  and  $p \in (n/(n-1), \infty)$  if  $n = 2$ . Moreover, set  $q = np/(n+p)$ . Then, (5) holds for*

$$C_p = |\Omega|^{\frac{2-q}{2q}} T_p,$$

where  $T_p$  is the constant in (24).

*Proof.* By zero extension outside  $\Omega$ , we may regard  $u \in H_0^1(\Omega)$  as an element  $u \in W^{1,q}(\mathbb{R}^n)$ ; note that  $q \leq 2$ , and  $q < 2$  if  $n = 2$ . Therefore, from Theorem A.1,

$$\|u\|_{L^p(\Omega)} \leq T_p \left( \int_{\Omega} |\nabla u(x)|_2^q dx \right)^{\frac{1}{q}}. \quad (25)$$

Hölder's inequality gives

$$\begin{aligned} \int_{\Omega} |\nabla u(x)|_2^q dx &\leq \left( \int_{\Omega} |\nabla u(x)|_2^{q \cdot \frac{2}{2-q}} dx \right)^{\frac{q}{2}} \left( \int_{\Omega} 1^{\frac{2}{2-q}} dx \right)^{\frac{2-q}{2}} \\ &= |\Omega|^{\frac{2-q}{2}} \left( \int_{\Omega} |\nabla u(x)|_2^2 dx \right)^{\frac{q}{2}}, \end{aligned}$$

that is,

$$\left( \int_{\mathbb{R}^n} |\nabla u(x)|_2^q dx \right)^{\frac{1}{q}} \leq |\Omega|^{\frac{2-q}{2q}} \|\nabla u\|_{L^2(\Omega)}, \quad (26)$$

where  $|\Omega|$  is the measure of  $\Omega$ . From (25) and (26), it follows that

$$\|u\|_{L^p(\Omega)} \leq |\Omega|^{\frac{2-q}{2q}} T_p \|\nabla u\|_{L^2(\Omega)}.$$

□

## B Power series arithmetic

Two types of Power Series Arithmetic (called Type-I PSA and Type-II PSA) were proposed by Kashiwagi [4, 5], and have been packaged in [6]. Both PSAs were originally designed to perform operations for sets of continuous functions defined on a closed interval  $D = [\underline{d}, \bar{d}]$  with  $\underline{d}, \bar{d} \in \mathbb{R}$ , written in the form

$$[u(x)] = \sum_{i=0}^n u_i x^i := \left\{ v \in C(D) : v(x) \in \sum_{i=0}^n u_i x^i \quad \forall x \in D \right\}. \quad (27)$$

where each  $u_i$  ( $i = 0, 1, 2, \dots, n$ ) is a real number or a real interval  $[u_i, \bar{u}_i]$ ,  $\underline{u}_i \leq \bar{u}_i$ . Type-I PSA performs such operations with neglecting terms of degree higher than  $n$ . Therefore, Type-I PSA gives approximate results of the operations. On the other hand, Type-II PSA gives a verified result of such operations, that is, the operation result from Type-II PSA includes the correct operation result in a strict mathematical sense. In this section, we introduce the original Type-II PSA in the one-dimensional case together with some operation examples. Subsequently, we present a generalization of Type-II PSA to the higher-dimensional cases in order to obtain a verified inclusion such as (21).

## B.1 Type-II PSA in the one-dimensional case

We consider a verified operation method for a set of continuous functions, written in the form (27). The addition operation and the subtraction operation are respectively performed as

$$[u(x)] + [v(x)] = \sum_{i=0}^n (u_i + v_i)x^i,$$

and

$$[u(x)] - [v(x)] = \sum_{i=0}^n (u_i - v_i)x^i.$$

The multiplication operation is performed as follows. We first multiply  $[u(x)]$  and  $[v(x)]$  without degree omissions:

$$[u(x)] \times [v(x)] = \sum_{i=0}^{2n} w_i x^i, \quad w_k = \sum_{i=\max(0, k-n)}^{\min(k, n)} u_i v_{k-i}.$$

Then, we reduce its degree from  $2n$  to  $n$  on the basis of the degree reduction defined as follows.

**Definition B.1** (Degree reduction). *For a power series  $[u(x)] = u_0 + u_1x + \dots + u_mx^m$  over  $D$ , the degree reduction  $[v(x)]$  to  $n$  ( $n < m$ ) is defined by*

$$[v(x)] = \sum_{i=0}^n v_i x^i,$$

where

$$v_i = u_i \quad (i = 0, 1, \dots, n-1) \quad \text{and} \quad v_n = \left\{ \sum_{i=n}^m u_i x^{i-n} \mid x \in D \right\}.$$

Thus, the terms of degree more than  $n$  are resorbed in the term of degree  $n$ . Therefore, the result of the multiplication by Type-II PSA includes the correct multiplication result.

**Remark B.2.** *When computing*

$$\left\{ \sum_{i=n}^m u_i x^{i-n} \mid x \in D \right\},$$

*one has to evaluate the range of the polynomial  $u_n + u_{n+1}x + \dots + u_mx^{m-n}$ . Since the common interval arithmetic occasionally over-estimates the range, one should use a method that gives the range more accurately, e.g., the Horner scheme, in order to obtain a precise multiplication result.*

We then apply Type-II PSA to  $C^\infty$ -functions (e.g,  $\log(\cdot)$  and  $\sin(\cdot)$ ) on the basis of the Taylor expansion with a remainder term. For a  $C^\infty$ -function  $f$ ,  $f(u_0 + u_1x + \dots + u_nx^n)$  is computed as

$$\begin{aligned} & f(u_0 + u_1x + \dots + u_nx^n) \\ & \subset f(u_0) + \sum_{i=1}^{n-1} \frac{1}{i!} f^{(i)}(u_0) (u_1x + \dots + u_nx^n)^i \end{aligned}$$



$$+ \frac{1}{n!} f^{(n)} \left( \text{hull} \left( u_0, \left\{ \sum_{i=0}^n u_i x^i \mid x \in D \right\} \right) \right) (u_1 x + \cdots + u_n x^n)^n, \quad (28)$$

by Taylor expanding  $f$  around  $u_0$ , where  $\text{hull}(a, b)$  denotes the convex hull of real numbers or real intervals  $a$  and  $b$ . Here, additions, subtractions, and multiplications in the above process are operated by Type-II PSA defined so far, and the expression

$$\left\{ \sum_{i=0}^n u_i x^i \mid x \in D \right\}$$

is similarly computed as mentioned in Remark B.2. The division can be operated as  $[u]/[v] := [u] \times f([v])$  with  $f(x) = 1/x$ , using the above method.

**Remark B.3.** *In our examples, the interval  $D$  in (28) contains zero in all cases; indeed, in the integration procedures described in Section 3, the domains  $S_{0,1}$ ,  $S_{1,0}$ , and  $S_{0,0}$  of integrations are translated to contain  $(0, 0)$  (Type-II PSA in the two-dimensional case will be introduced in B.3 using the one-dimensional method). Hence, in our examples,*

$$\text{hull} \left( u_0, \left\{ \sum_{i=0}^n u_i x^i \mid x \in D \right\} \right) = \left\{ \sum_{i=0}^n u_i x^i \mid x \in D \right\}$$

*always holds in (28).*

**Remark B.4.** *Basically, Type II-PSA is designed to ensure that the coefficients of degree less than  $n$  are points (real numbers), and the coefficient of degree  $n$  is a real interval. However, in an actual computation, in order to strictly verify all results from Type II-PSA, the coefficients of degree less than  $n$  are often intervals that arise only from rounding error enclosures.*

## B.2 Examples of Type-II PSA

Here, we present simple examples of Type-II PSA where (degree of PSA  $n$ ) = 2 and  $D = [0, 0.1]$ ,  $[u(x)] = 1 + 2x - 3x^2$  and  $[v(x)] = 1 - x + x^2$ .

The addition operation and the subtraction operation are respectively performed as  $[u(x)] + [v(x)] = 2 + x - 2x^2$  and  $[u(x)] - [v(x)] = 0 + 3x - 4x^2$ .

The multiplication operation is performed as follows. We first multiply them as

$$\begin{aligned} [u(x)] \times [v(x)] &= 1 + x - 4x^2 + 5x^3 - 3x^4 \\ &= 1 + x + (-4 + 5x - 3x^2)x^2 \end{aligned}$$

without degree omissions. Since interval arithmetic gives

$$-4 + (5 - 3 \times [0, 0.1]) \times [0, 0.1] \subset -4 + [0, 0.5] \subset [-4, -3.5],$$

we determine the multiplication result as

$$[u(x)] \times [v(x)] = 1 + x - [-4, -3.5]x^2,$$

on the basis of the the degree reduction in Definition B.1.

The computation  $\log([u(x)])$  is performed as follows. The range of  $[u(x)]$  is computed as

$$1 + (2 - 3 \times [0, 0.1]) \times [0, 0.1] \subset 1 + [0, 0.2] \subset [1, 1.2].$$

We then compute the second degree Taylor expansion, with a remainder term, of  $\log(t)$  around 1 (the constant term of  $[u(x)]$ ) on  $[1, 1.2]$  as

$$0 + (t - 1) - \frac{1}{2[1, 1.2]^2}(t - 1)^2.$$

By substituting  $[u(x)]$  for  $t$  in this expansion, we have

$$0 + (2x - 3x^2) - \frac{1}{2[1, 1.2]^2}(2x - 3x^2)^2.$$

Consequently, by reducing this expression using Type-II PSA, we have

$$\log([u(x)]) = 0 + 2x + \left[-5, -\frac{143}{36}\right] x^2.$$

### B.3 Type-II PSA in the higher-dimensional cases

One-dimensional Type-II PSA is designed for power series that have real or real interval coefficients. In fact, the set of coefficients in Type-II PSA can be generalized to any set equipped with the four arithmetic operations. Moreover, the set of power series is endowed with the four arithmetic operations by Type-II PSA. Therefore, Type-II PSA can be generalized to two-dimensional cases by replacing its coefficients with one-dimensional power series. To be precise, by replacing each coefficient  $u_i$  in

$$[u(x)] = \sum_{i=0}^n u_i x^i, \quad x \in D_u \tag{29}$$

with one-dimensional power series

$$[v_i(y)] = \sum_{j=0}^n v_{i,j} y^j, \quad y \in D_v,$$

we can regard  $[u]$  as a two-dimensional power series

$$[u(x, y)] = \sum_{i=0}^n [v_i(y)] x^i = \sum_{i=0}^n \sum_{j=0}^n v_{i,j} x^i y^j, \quad (x, y) \in D_u \times D_v. \tag{30}$$

Thus, Type-II PSA for the one-dimensional case is naturally carried over to the two-dimensional case. In the same way, Type-II PSA can be also applied to higher-dimensional cases, that is, by replacing each coefficient  $u_i$  in (29) with  $n$ -dimensional power series,  $(n + 1)$ -dimensional power series with the four arithmetic operations are defined.

## References

- [1] Thierry Aubin. Problèmes isopérimétriques et espaces de Sobolev. *Journal of Differential Geometry*, 11(4):573–598, 1976.
- [2] Henning Behnke. The calculation of guaranteed bounds for eigenvalues using complementary variational principles. *Computing*, 47(1):11–27, 1991.
- [3] Pierre Grisvard. *Elliptic problems in nonsmooth domains*, volume 69. SIAM, 2011.

- [4] Masahide Kashiwagi. Power series arithmetic and its application to numerical validation. In *Proc. 1995 International Symposium on Nonlinear Theory and its Applications (NOLTA 95 Symposium)*, pages 251–254, Las Vegas, U.S.A., 1995.
- [5] Masahide Kashiwagi. Power series arithmetic and its application to numerical validation. *The Institute of Electronics, Information and Communication Engineers*, 95(296):1–8, 1995.
- [6] Masahide Kashiwagi. KV library, 2015. <http://verifiedby.me/kv/>.
- [7] Xuefeng Liu. A framework of verified eigenvalue bounds for self-adjoint differential operators. *Applied Mathematics and Computation*, 267:341–355, 2015.
- [8] Shinya Miyajima. Numerical enclosure for each eigenvalue in generalized eigenvalue problem. *Journal of Computational and Applied Mathematics*, 236(9):2545–2552, 2012.
- [9] Ramon E Moore, R Baker Kearfott, and Michael J Cloud. *Introduction to interval analysis*. Siam, 2009.
- [10] Mitsuhiro T Nakao. A numerical approach to the proof of existence of solutions for elliptic problems. *Japan Journal of Applied Mathematics*, 5(2):313–332, 1988.
- [11] Mitsuhiro T Nakao. Numerical verification methods for solutions of ordinary and partial differential equations. *Numerical Functional Analysis and Optimization*, 22(3-4):321–356, 2001.
- [12] Mitsuhiro T Nakao and Yoshitaka Watanabe. Numerical verification methods for solutions of semilinear elliptic boundary value problems. *Nonlinear Theory and Its Applications, IEICE*, 2(1):2–31, 2011.
- [13] Michael Plum. Computer-assisted existence proofs for two-point boundary value problems. *Computing*, 46(1):19–34, 1991.
- [14] Michael Plum. Explicit  $H^2$ -estimates and pointwise bounds for solutions of second-order elliptic boundary value problems. *Journal of Mathematical Analysis and Applications*, 165(1):36–61, 1992.
- [15] Michael Plum. Computer-assisted enclosure methods for elliptic differential equations. *Linear Algebra and its Applications*, 324(1):147–187, 2001.
- [16] Michael Plum. Existence and multiplicity proofs for semilinear elliptic boundary value problems by computer assistance. *Jahresbericht der Deutschen Mathematiker Vereinigung*, 110(1):19–54, 2008.
- [17] Michael Plum. Computer-assisted proofs for semilinear elliptic boundary value problems. *Japan journal of industrial and applied mathematics*, 26(2-3):419–442, 2009.
- [18] S.M. Rump. INTLAB - INTerval LABoratory. In Tibor Csendes, editor, *Developments in Reliable Computing*, pages 77–104. Kluwer Academic Publishers, Dordrecht, 1999. <http://www.ti3.tuhh.de/rump/>.
- [19] Ulrike Storck. Numerical integration in two dimensions with automatic result verification. *Mathematics in science and engineering*, 189:187–224, 1993.
- [20] Akitoshi Takayasu, Xuefeng Liu, and Shin’ichi Oishi. Remarks on computable a priori error estimates for finite element solutions of elliptic problems. *Nonlinear Theory and Its Applications, IEICE*, 5(1):53–63, 2014.

- [21] Giorgio Talenti. Best constant in Sobolev inequality. *Annali di Matematica pura ed Applicata*, 110(1):353–372, 1976.
- [22] Kazuaki Tanaka, Kouta Sekine, Makoto Mizuguchi, and Shin’ichi Oishi. Sharp numerical inclusion of the best constant for embedding  $H_0^1(\Omega) \hookrightarrow L^p(\Omega)$  on bounded convex domain. *to appear*.
- [23] Kazuaki Tanaka, Kouta Sekine, Makoto Mizuguchi, and Shin’ichi Oishi. Numerical verification of positiveness for solutions to semilinear elliptic problems. *JSIAM Letters*, 7:73–76, 2015.
- [24] Kazuaki Tanaka, Akitoshi Takayasu, Xuefeng Liu, and Shin’ichi Oishi. Verified norm estimation for the inverse of linear elliptic operators using eigenvalue evaluation. *Japan Journal of Industrial and Applied Mathematics*, 31(3):665–679, 2014.